

Weekly Report

梅鸿辉

January 21, 2018

1. 分辨率自适应可视化

1.1 目前和未来计划进行的尝试

已有工作 可能工作 具体布置 问题探究 时间安排

散点图测试系统 | 测试结果上传服务器整合¹²¹

点半径 \times 图长宽拟合 更多参数 (最小点间距) 模仿之前测试进行更多试验¹²³

感知顺序 限时测试 比较差异? 寻找相关工作¹²⁴

不同数据分布 尝试一下

- 密度
- 分布区域 (中心 vs 四周)

= 最小点间距等因素的出现区域

点半径本身有编码

Scagnostics¹²⁵ 挑选部分写函数计算 随机生成测试数据, 然后再找相关性?

反过来 (给定圆半径求其它)

Figure 1: 目前和未来计划进行的尝试

1.2 数据科学入门

简单的系统性学习了一下概率论和假设检验方面的内容 (Figure 3)

1.3 Scagnostics

考虑用scagnostics的一些参数作为定量分析的基准 (模型你和函数的横轴), 进行了一些简单的学习和整理 (Figure 2), 准备进行一些初步的实现。

Scagnostics 整理	Graph-theoretic scagnostics
convex hull / nonconvex hull / minimum spanning tree (MST)	Convex Hull Alpha Hull MST
- outliers (quantiles 相关 $W = q_{10} + 1.5(q_{95} - q_{10})$)	
$C_{\text{outlying}} = \text{length}(T_{\text{outliers}}) / \text{length}(T)$	H-convex hull
- shape	A: alpha hull
$C_{\text{convex}} = \text{area}(A) / \text{area}(H)$	T: MST
$C_{\text{slangy}} = 1 - \sqrt{4\pi \text{area}(A)} / \text{perimeter}(A)$	
$C_{\text{stringy}} = \text{diameter}(T) / \text{length}(T)$	
$C_{\text{straight}} = d(\text{st}(t_j, t_k)) / \text{diameter}(T)$ (t_j, t_k : 直径两端)	
- trend	
$C_{\text{monotonic}} = r^2_{\text{poorman}}$	
- density	
$C_{\text{skew}} = (q_{90} - q_{50}) / (q_{10} - q_{50})$ / skewness of histogram of MST edges lengths	The use of minimal spanning tree to characterize the 2D cluster galaxy distribution
$C_{\text{clumpy}}(T) = \max_j [1 - \max_k [\text{length}(e_k) / \text{length}(e_j)]]$ (e_k : runt set of e_j)	The RUNT Test for Multimodality
- coherence	
$C_{\text{clustrate}} = \frac{1}{ V(D) } \sum_{v \in V(D)} \cos \theta_{v, \text{nearest}} $ ($V^{(2)}$: vertices of degree 2)	

Figure 2: 一些简单的scagnostics整理

2. Others

- 理头
- 期末课题报告（计算机安全）

Papaer Reading

peception

[1] W. K. Estes and H. A. Taylor, "Visual detection in relation to display size and redundancy of critical elements," *Percept. Psychophys.*, vol. 1, no. 1, pp. 9–16, 1966.

标题有display size但内容更多的和element number相关。

[2] J. Palmer, "Attentional limits on the perception and memory of visual information.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 16, no. 2, pp. 332–350, 1990.

没太看懂。

[3] R. Klein, "Inhibitory tagging system facilitates visual search," *Nature*, vol. 334, no. 6181, pp. 430–431, 1988.

重点在tagging，没太大关联。

[4] M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri, “Perception of average value in multiclass scatterplots,” *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2316–2325, 2013.

主要是多类，没什么关联。

[5] M. Bayer, M. Burgholzer, H. Hörl, and A. Zobl, “Retinal Coding and Mapping,” *Inf. Vis.*, 2010.

CS专业的课程作业，没什么用。

[6] J. Li, J. J. Van Wijk, and J. B. Martens, “Evaluation of symbol contrast in scatterplots,” *IEEE Pacific Vis. Symp. PacificVis 2009 - Proc.*, pp. 97–104, 2009.

之前两篇基本一样的文章的前置文章，类似的结构，研究的较为粗泛。

[7] W. S. Cleveland and R. McGill, “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods,” *J. Am. Stat. Assoc.*, vol. 79, no. 387, pp. 531–554, 1984.

统计不同视觉通道感知的准确性等的差异。

[8] [PPT]Graphical Perception.

讲述了一些基础理论，挺有用的。

[9] N. Kong, *Techniques for Modifying and Augmenting Existing Charts for Improved Usability*. 2013.

似乎是逆向工程相关，没有太多感知的内容。

[10] W. S. Cleveland and R. McGill, “Graphical Perception and Graphical Methods for Analyzing Scientific Data,” *Am. Assoc. Adv. Sci.*, vol. 229, no. 4716, pp. 828–833, 1985.

看起来Cleveland的著作是很有必要读一下，希望能早点寄到吧。

scagnostics

[1] G. Scagnostics, “Graph-theoretic scagnostics,” pp. 157–164, 2005.

主要是这篇，下面两篇都是这篇里的引用去看具体算法的。

[2] J. A. Hartigan and S. Mohanty, “The RUNT Test for Multimodality,” *J. Classif. Classif. Lit. Autom. Search Serv. / plus Cl.*, vol. 9, no. 1, pp. 63–70, 1992.

[3] C. Adami and A. Mazure, “The use of minimal spanning tree to characterize the 2D cluster galaxy distribution,” *Astron. Astrophys. Suppl. Ser.*, vol. 134, no. 2, p. 9, 1998.

Works Progresses

TASK	PROGRESS	TODO	ISSUES	DATE
分辨率自适应可视化	制作了简单的测试系统	学习/咨询相关理论基础		
专利（两个）	1/2	继续按照要求修改		
VisComposer	投PFAST	等待回复		
VisEvo		idea evaluation		下周
JVLC	publicated			
电子学报	已提交	等待回复		
ECharts论文	投PFAST	等待回复		
IDEA项目	1/10 展板			

统计学

描述数据集

数据本身 → 统计直方图

单个 $len / max / min \dots \rightarrow mean / median (quantile) / mode$

离散度: $range / variance (standard deviation) / 25/75 \text{ quantile}$

相关: $covariance / correlation / outlier$

辛普森悖论

统计学教材

OpenIntro Statistics <https://www.openintro.org/stat/textbook.php>

OpenStax Introductory Statistics <https://openstax.org/details/introductory-statistics>

概率

独立, 条件概率 $P(E, F) = P(E|F)P(F)$

贝叶斯定理 $P(E|F) = P(F|E)P(E) / [P(F|E)P(E) + P(F|\neg E)P(\neg E)]$

随机变量, 离散分布, 连续分布, 概率密度函数 → 累计分布函数

正态分布, 中心极限定理

概率论教材

introduction to probability https://www.dartmouth.edu/~chance/teaching_aids/books_article

假设与推断

假设与假设检验

零假设 H_0 , 替代假设 H_1 ; 统计 → 拒绝 H_0 ?

显著性: $p(\text{拒绝 } H_0 \text{ (正确)}) / \text{势: } 1 - p(\text{没有拒绝 } H_0 \text{ (错误)}) / \text{单边检验} / p \text{ 值}$

置信区间

5

p -hacking → 贝叶斯推断

延伸学习: 上述“统计学教材”; 公开课 <https://www.coursera.org/specializations/statistics>